

len_distribution_repseqs_asvs_cutadaptq20minlen150

May 7, 2024

1 Length distribution of ASV representative sequences from Dada2 after Cutadapt

Plotting the distribution of sequence length for the representative sequences of dada2, after cutadapt (minimum quality 20 and minimum length 150 of forward reads).

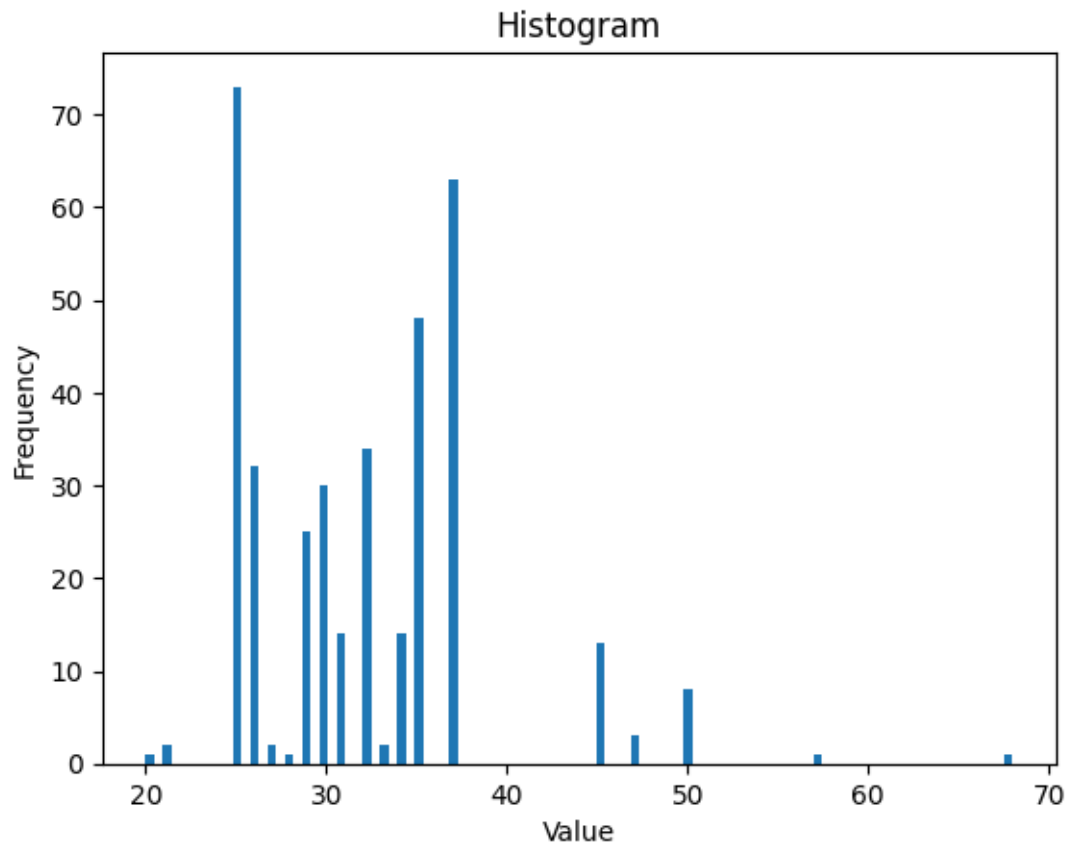
```
[ ]: import pandas as pd

#file_path = '/media/renato/SSD1TB/Projects/UGA_RACS/16S/Qiime2/dada2/q20_150bp/
↳rep-seqs-paired-end_wallace2018_assingle_forward_q20_150bp-dada2/
↳freq_asv_lenghts_q20_150bp.tsv'
file_path = '/media/rsantos/4TB_drive/Projects/UGA_RACS/16S/Qiime2/dada2/
↳q20_150bp/rep-seqs-paired-end_wallace2018_assingle_forward_q20_150bp-dada2/
↳freq_asv_lenghts_q20_150bp.tsv'
repseqs_q20_150bp_df = pd.read_csv(file_path, sep='\t')
repseqs_q20_150bp_df.head()
```

```
[ ]:   Freq  Length
0      3     150
1      2     151
2      3     152
3      2     153
4      5     155
```

```
[ ]: filtered_lt100_df = repseqs_q20_150bp_df[repseqs_q20_150bp_df['Length'] < 100]
filtered_gt150_df = repseqs_q20_150bp_df[repseqs_q20_150bp_df['Length'] > 150]
filtered_gt190_df = repseqs_q20_150bp_df[repseqs_q20_150bp_df['Length'] > 190]
```

```
[ ]: import matplotlib.pyplot as plt
values = filtered_lt100_df['Length']
frequencies = filtered_lt100_df['Freq']
plt.hist(values, weights=frequencies, bins=100)
plt.xlabel('Value')
plt.ylabel('Frequency')
plt.title('Histogram')
plt.show()
```



```
[ ]: import matplotlib.pyplot as plt
values = filtered_gt190_df['Length']
frequencies = filtered_gt190_df['Freq']
plt.hist(values, weights=frequencies, bins=100)
plt.xlabel('Value')
plt.ylabel('Frequency')
plt.title('Histogram')
plt.show()
```

